DOCUMENT RESUME

ED 275 707 TM 860 617

AUTHOR Strum, Irene
TITLE Determining Program Effectiveness in Special Education.
PUB DATE Apr 86
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Accountability; Educational Testing; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluation Problems; Individualized Education Programs; *Program Effectiveness; *Program Evaluation; School Surveys; *Special Education

ABSTRACT

Discussions were held with administrative staff and evaluators from the New York City Public Schools in order to identify sources of data for evaluating special education programs. Data which are generally available include: (1) citywide reading test scores; (2) number of diplomas issued; (3) number of students decertified; (4) number of students moved to less restrictive environments; (5) attendance; (6) number of annual promotions; (7) number of students beginning high school diploma programs; (8) number of dropouts in high school age group; (9) test scores; and (10) amount of academic mainstreaming. Use of the Individualized Education Program in conducting evaluations is not recommended at this time. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275707

DETERMINING PROGRAM EFFECTIVENESS IN SPECIAL EDUCATION

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

I. Strum

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Irene Strum
New York City Public Schools
Division of Special Education
Office of Program Development
Testing Unit

TM 860 617

Paper delivered at the 1986 American Educational Research Association Annual Meeting, April, 1986

Summary

1. Citywide reading test scores by program and grade
2. Number of diplomas issued (by program)
3. Number of students decertified by grade and program
4. Number of students moved to less restrictive environments by grade and program
5. Attendance by grade (or school level) and program
6. Number of annual promotions by grade in SIS or resource rooms
7. Numbers of students beginning high school age programs in diploma and nondiploma tracks
8. Number of dropouts by program in high school age programs
9. Test scores from assessment: 3-year comparisons by programs and age/grade
10. Amount of academic mainstreaming by program and school level

Determining Program Effectiveness In Special Education

Program evaluation in special education is a relatively new concept. Historically, the researcher carried out a case study to determine the effectiveness of an intervention for one student (MacLeod, Andrews, & Grove, 1980); this approach was based upon the uniqueness of each individual student in special education. Conferences held by professional educational organizations have included sessions on this subject since 1982, but few responses have been presented that may be implemented by schools and school districts. In 1982, the American Educational Research Association (AERA) had several symposia on the topic at its annual meeting; in a presentation on the state of the art in a school district, Farber (1982) reported on surveys and needs assessments. During this same year a symposium on the National Council for Measurement in Education (NCME) program on a model for special education program evaluation was given by New York City special educators and evaluators (Strum, 1982). The 1984 programs of AERA and the Council for Exceptional Children (CEC) listed several sessions on the same topic.

Since the 1981-1982 school year, California has been implementing program evaluation in special education (Scheffelin, Ragsdale, & Martinez, 1985) which looks at "satisfaction with special education services; criteria for eligibility; student performance, including differential graduation standards; and program costs (p 87)."

The Educational Leaders Consortium (ELC) (Elliott & Hall, 1985) identified as indicators of student and school performance student attendance, achievement, and dropouts; teacher attendance, verbal ability, job satisfaction; class size, instructional processes, climate; community participation. Working with the Chief State School Officers, the ELC raised the following questions which are relevant to our inquiry: "Are student achievement tests measuring what is taught or what should be taught? Are the measures ones that have a known or highly likely relationship with student performance? Is there a distinction between measures of factors that educators or policymakers can change and those they cannot? Are the measures understandable? (p.8,9)".

Purpose of Evaluation

The decision-maker must determine the purpose(s) for evaluating program effectiveness before selecting the criteria to implement an investigation. Is it accountability? Do we need data for administrators to plan program changes and staff development? Are we concerned with evaluating the instructional program for individual students or specific groups of students?

The research related to the first two questions is meager. In general, there are no positive suggestions for the practitioner. In funded program evaluation there have been two models: (a) criterion referenced (b) norm referenced.

Models of Evaluation

In the criterion referenced model, aggregate information was obtained on the kinds of objectives students mastered, the number of objectives students mastered, and the percentage of mastery for individual students Major weaknesses of this approach include inequality of objectives and fragmentation of the curriculum.

In the norm referenced model, pre- and posttesting is implemented and the appropriate formula used to determine change; presently the model compares NCEs, which are normalized percentiles to determine group growth Weakness of this approach includes the generally low functional level of the students which tends to make scores unreliable; however, with the adoption of new eligibility guidelines, the lowest achievers are generally not included in the population.

Much research has been written on evaluating individual student programs, most of which is related to IEPs. The most positive approach is probably the Deno and Merkin model which emphasizes the frequent use of curriculum related measures. These measures are reliable since they consist of many samples of behavior over time. They are valid because they are directly related to instruction and are considered to have content or representational validity (Shinn, 1985). In a CEC session, Shinn also indicated that the measure of student performance must be based on the general education curriculum. In St. Cloud, Minnesota, the measurement was based on how well the student read the average level of a basal reader in a specific grade for a period of three days. The problem with using curriculum based assessment is the difficulty of aggregating data (Shinn, 1985). Tindal (1985) supported the use of the same measures for determining program effectiveness and indicated inadequacy of grade equivalents. Hammill (1985) suggested the use of mastery tests.

Procedures in the Local Educational Agency (LEA)

Discussions were held with four Deputy Assistant Superintendents and/or designated staff during the 1984-85 school year on the subject of criteria for evaluating program effectiveness. It was also the subject of discussion with the superintendent and designated staff of citywide programs. A meeting was also held with a representative from the Chancellor's office on evaluation. In addition, the topic was discussed at a meeting of the Instructional Testing Committee, composed of representatives of regional and citywide programs. Data that are generally available were discussed.

Results: Criteria for Evaluating Program Effectiveness

The criteria which were identified resulted from the procedures described above. The utility of each criterion is addressed along with strengths and weaknesses so that a rationale for justifying use or nonuse of a criterion is developed.

Citywide reading test scores by program and grade. The use of these scores parallels general education and would permit educators to look at special education in a manner similar to general education. The scores are a means of accountability and permit gross decisions for program change and staff development. Their use would require the collection of baseline data in order to determine expectations and "standards". However, not all students have citywide reading test scores. Kindergarteners and first graders have not been tested in the citywide program as yet; high school students are tested in another program.

Moreover, 13% of students in special education are exempt from city and state examinations and a different system would be needed for program evaluation for these students. Arguments are sometimes given on the unreliability of scores because so many students function seriously below grade level; however, the Office of Educational Assessment (OEA) preliminary research found support for reliability on the basis of a progression of year-by-year growth through the grades (Tobias, 1984). Modifications are frequently argued as making scores uninterpretable; on the other hand, modifications provide for more reliable student measurement and attempt to control for handicapping condition.

Number of diplomas issued (by program). Diploma data are clearly useful as accountability statistics. The bottom line for handicapped students is their ability to obtain jobs. Employability is maximized by possession of a diploma. Related to this are numbers of students by program meeting State Reference Points or minimal scores in New York State testing programs (PEP, PCT, RCT). By itself, numbers of diplomas cannot direct meaningful change; however, the addition of state test information, presently in reading, writing, and math, and within the next few years, in science, second language, the social studies as well, should facilitate direction for program change and staff development. Student goals and objectives may also be affected, since failure to meet minimal requirements necessitates remediation, even in special education.

Number of students decertified by grade and program. These data may also be used for accountability. They are bottomline data meeting one of the goals for mildly and moderately handicapped students. However, these data are not sufficient for school system accountability without additonal information on whether such students earned diplomas in general education.

Number of students moved to less restrictive environments by grade and program. These data are similar to the above. They are process evaluation information which may be useful for indepth investigation of programs at different gradel levels and for reviewing IEP recommendations for individual students.

Attendance by grade (or school level) and program. Attendance is viewed as attitudinal information; change in attitude frequently precedes cognitive change. These data are used for accountability in process evaluation.

Number of annual promotions by grade for SIS or resource room students. These data along with decertifications and diplomas also provide information for accountability. They should assist decision-makers in projecting the number of diplomas in the future.

Numbers of students beginning high school age programs in diploma and nondiploma tracks. These data may also be used for accountability of pre-secondary school programs and to assist decision-makers in advocating program changes and staff development.

Number of dropouts by program in high school age programs. These data provide accountability information and may be useful for recommending program change.

Test scores from assessment: 3-year comparsions by program and age/grade. These data are useful for individual student program planning. Aggregates across districts, regions, and total division may provide useful data for determining effectiveness of special education interventions.

Amount of academic mainstreaming by program and school level. These process evaluation data are useful in determining effectiveness of efforts toward less restrictive environments.

Note that the use of the IEP was omitted. No one believed that IEP information would be useful at this time because of the lack of standardization of goal and objective statements, the lack of systematic evaluation procedures for determining mastery, and the inequality of goals and objectives in relation to scope, depth, difficulty, etc. Moreover, the new Part 100 Regulations (University of State of New York) which focus on equal access for the handicapped emphasized the need for identifying criteria more closely related to general education criteria.

Recommendations

The criteria which were identified during discussion with field and central staff can be accessed under the present systems of data collection and used to provide information to decision-makers. Improving the IEP and research and development to improve goals, objectives, and evaluations of mastery should be undertaken to provide curriculum based information that will be specifically relevant to Special Education for planning program change and staff development activities.

References

Elliott, E.J., & Hall, R. (1985) Indicators of performance: Measuring the educators. Educational Measurement: Issues and Practices, Summer, 6-9.

Farber, I.J. (1982) Evaluation services for special education: The local education agency perspective. New York City: AERA

Hammill, D. (1985) Informal talk. Session B-1, CED annual conference, Anaheim, CA.

MacLeod, J.C., Andrews, J., & Grove, D.N. (1980) Single subject procedures. Seattle, WA: WESTAR.

Scheffelin, M. M., Ragsdale, D., & Martinez, J. (1985) Program evaluation in special education: From policy to practice in California. Journal of Learning Disabilities, 18, 87-88.

Shinn, M. (1985) Session T14, CEC annual conference, Anaheim, CA.

Strum, I. (1982) Symposium: A model for program evaluation in special education. NYC: NCME.

Tindal, G. (1985) Session T14, CEC annual conference, Anaheim, CA.

Tobias, R. (1984) Personal communication.

University of the State of New York. Part 100 regulations and special education. Albany, NY: State Education Department, Office for Education of Children with Handicapping Conditions.